Image Segmentation in Complex Backgrounds using an Improved Generative Adversarial Network

Mei Wang, Yiru Zhang*
School of Information Science and Engineering, Shandong Agriculture and Engineering University, Jinan 250100, China[1]
Jinan Intellectual Property Protection Centery, Jinan 250100, Shandong, China[2]

***Abstract*—As technology advances, solving image segmentation challenges in complex backgrounds has become a key issue across various fields. Traditional image segmentation methods underperform in addressing these challenges, and existing generative adversarial networks (GANs) also face several problems when applied in complex environments, such as low generation quality and unstable model training. To address these issues, this study introduces an improved GAN approach for image segmentation in complex backgrounds. This method encompasses preprocessing of complex background image datasets, feature reduction encoding based on cerebellar neural networks, image data augmentation in complex backgrounds, and the application of an improved GAN. In this paper, new generator and discriminator network structures are designed and image data enhancement is implemented through self-play learning. Experimental results demonstrate significant improvements in image segmentation tasks in various complex backgrounds, enhancing the accuracy and robustness of segmentation. This research offers new insights and methodologies for image processing in complex backgrounds, holding substantial theoretical and practical significance.**

***Keywords*—Generative Adversarial Networks (GANs); complex backgrounds; image segmentation; data augmentation; feature dimensionality reduction encoding**

I. Introduction

As technology progresses, image processing technology has been widely applied, especially in the field of image segmentation in complex backgrounds, which is a key technology in many areas including medical imaging, intelligent video surveillance, and machine vision [1-17]. However, image segmentation in complex backgrounds remains a challenging research problem, as features in images may become difficult to recognize due to factors such as lighting, texture, and color in complicated environments [18-20].

The development of image segmentation technology plays a crucial role in improving the accuracy and efficiency of image recognition. Using improved GANs for image segmentation in complex backgrounds holds a wide application prospect and significant research value [21-24]. This technology can enhance the accuracy and efficiency of image segmentation and can effectively enhance and reduce the dimensionality of images, thereby increasing the flexibility and practicality of image processing [25-28].

Existing work on image segmentation in complex backgrounds typically employs traditional image processing and machine learning methods. These include pixel-level feature extraction and classification techniques such as edge detection and region growing; traditional machine learning algorithms like Support Vector Machines and Random Forests are also used for image segmentation. Additionally, some existing work has explored deep learning technologies, such as Convolutional Neural Networks (CNN) or Fully Convolutional Networks (FCN), to enhance image segmentation performance. However, these existing methods often face challenges and limitations when dealing with image segmentation in complex backgrounds. For example, traditional feature extraction methods may not adequately capture the semantic information in complex backgrounds, leading to inaccurate segmentation results; traditional machine learning algorithms may lack generalization capability over complex background datasets, making it difficult to adapt to changes in different scenes; and methods based on deep learning might require extensive labeled data and computational resources, and may be limited by the quality of data and the design of the model. However, existing research methods have many defects and shortcomings. On one hand, traditional image segmentation methods often fail to effectively process images in complex backgrounds, and cannot effectively recognize and segment key features of the images [29-32]. On the other hand, existing GANs may encounter issues such as low quality of generated images and unstable model training when dealing with image segmentation in complex backgrounds. This limits the breadth and depth of application of GANs in image segmentation under complex backgrounds [33-35].

Compared to existing work, this paper proposes a series of innovative methods in the field of image segmentation in complex backgrounds. Firstly, the paper utilizes feature reduction encoding technology based on cerebellar neural networks to effectively extract key features of images in complex backgrounds during the preprocessing stage, providing richer information for subsequent segmentation processes. Secondly, by introducing an improved GAN, new generator and discriminator network structures are designed, which perform better in handling image segmentation in complex backgrounds compared to traditional methods. Most importantly, the paper innovatively employs self-play learning to implement image data enhancement, thereby improving the model's adaptability and generalization performance in complex backgrounds.

II. Dataset Preprocessing and Feature Dimensionality Reduction Encoding

Fig. 1 shows the workflow diagram for the image segmentation model in complex backgrounds. For image

*Corresponding Author.

segmentation tasks, data processing during the preprocessing phase is crucial, as proper preprocessing can improve the efficiency of subsequent model training and the accuracy of the final segmentation results. In image segmentation in complex backgrounds, due to the complexity of the background, image features are influenced by factors such as lighting, texture, and color, making the feature distribution complex and varied. Therefore, it is necessary to transform and optimize the image data through preprocessing of complex background images to minimize these impacts and enhance model performance.

First, data logarithmization is an important preprocessing step. Since images in complex backgrounds may exhibit a wide range of grayscale or color differences, this can lead to excessively high image contrast, causing some important features to be overlooked in subsequent processing. Data logarithmization can reduce this gap, lowering the contrast while retaining important image features, which is very beneficial for subsequent feature dimensionality reduction encoding and image segmentation.

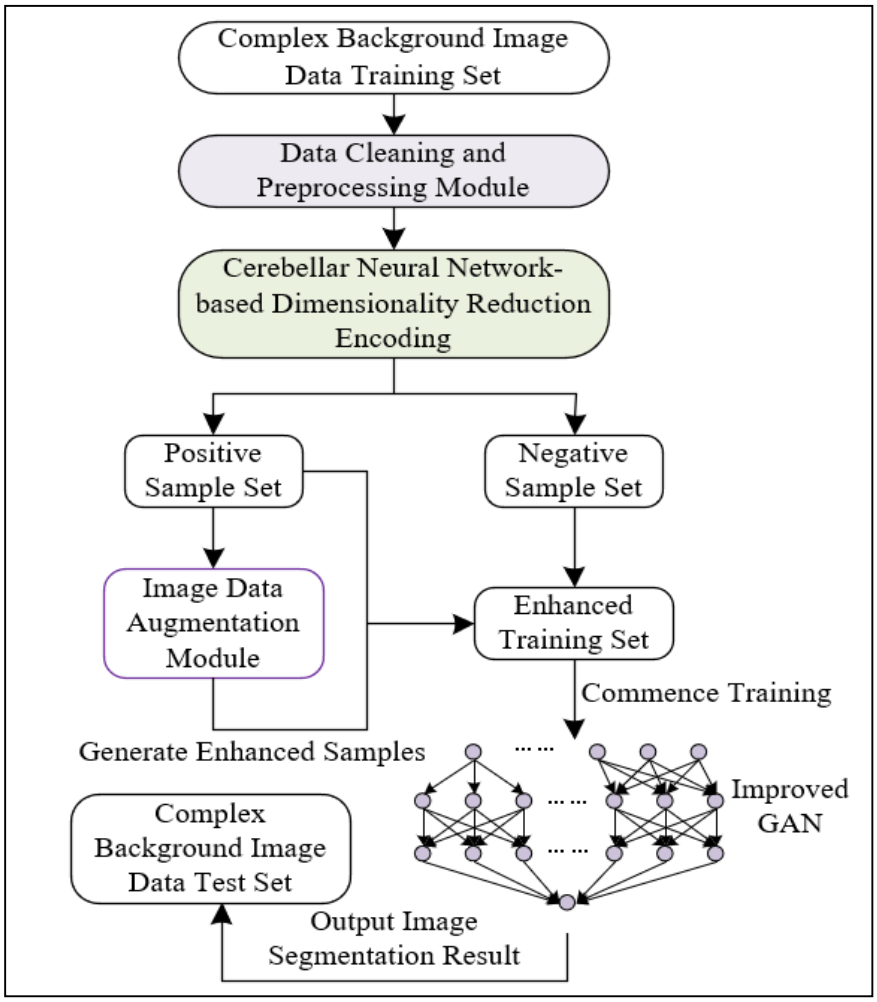


Fig. 1. Workflow for the image segmentation model in complex backgrounds.

Assuming the original data is represented by z, and the logarithmized data by $z^{(LOG)}$, the logarithmization processing formula is given by:

$$z^{(\mathrm{LOG})} = \ln\left(z+\varphi\right) \tag{1}$$

Secondly, data normalization is also an important part of complex background image preprocessing. Data normalization eliminates the differences in magnitude and scale of image data, allowing the data to be processed on the same scale, which is beneficial for improving the stability and performance of model training. For image segmentation using GANs, normalization of the data can increase the convergence speed of the model and reduce instability during model training.

Assuming the features of a sample after logarithmization are represented by $z^{(LOG)}$, where $z^{(LOG)}=(z_1^{(LOG)},z_2^{(LOG)},\ldots,z_u^{(LOG)},\ldots,z_l^{(LOG)})$, the standardized data is represented by $z^{(STA)}$, and the standard deviation $STA(z)$ is calculated using the following formulas:

$$STA(z)=\sqrt{\frac{1}{l-1}\sum_{u=1}^{l}\left(z_u^{(\log)}-\frac{1}{l}\sum_{k=1}^{l}z_k^{(\log)}\right)} \tag{2}$$

$$z^{(STA)}=\frac{z^{(\log)}}{STA(z)} \tag{3}$$

In scenarios of image segmentation in complex backgrounds, image data usually has high-dimensional features because it needs to contain sufficient information to describe the various possibilities of complex backgrounds. However, high-dimensional features also pose challenges for model training and optimization, such as high computational complexity and overfitting. Therefore, reducing the dimensionality of high-dimensional features, while retaining key information, is very important for enhancing model performance and the accuracy of image segmentation.

This paper chooses to use cerebellar neural networks for feature dimensionality reduction encoding. Possible reasons include the distributed storage and fault tolerance of cerebellar neural networks, making them ideal tools for feature dimensionality reduction encoding. In cerebellar neural networks, the design of the compressed storage part is key, as it determines whether the network can effectively reduce the dimensionality of high-dimensional features.

In the scenario of image segmentation in complex backgrounds, the goal of this paper is to effectively encode and store high-dimensional image features, to facilitate more efficient and accurate image segmentation in subsequent processing. To achieve this goal, specific designs of the compressed storage part of the cerebellar neural network have been utilized, mainly involving techniques such as linear shift registers and binary hashing mapping.

Initially, binary code strings are used to store virtual mapping addresses. This method converts high-dimensional image features into relatively low-dimensional binary strings, thereby reducing data complexity. Furthermore, linear shift registers are used for hashing mapping in cerebellar neural networks, and the results are stored. Linear shift registers are efficient tools for processing binary data, and hashing mapping can project high-dimensional data into a lower-dimensional space, further reducing data complexity. Fig. 2 displays the schematic of the working process of the linear shift register.

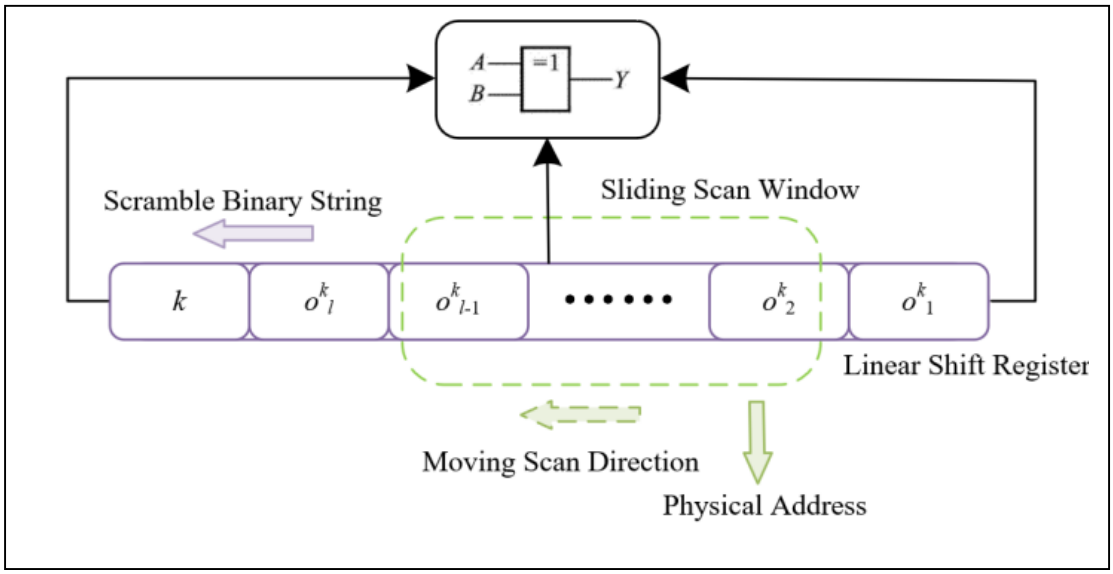


Fig. 2. Working process of the linear shift register.

Assuming the virtual associative mapping address of a sample is represented by $o=(o^1,o^2,...,o^k,...,o^v)$, where each o^k address occupies $n_c=[LOG_2([w/v])+1]+1$ binary bits, $k = 1, 2, ..., v$. For each o^k, $k=1,2,...,v$, there is a linear shift register performing the hash mapping. Assuming the storage address label uses n_{IN} binary bits, then $n_{IN}=[LOG_2([v])+1]+1$ ensures the storage requirements are met. Thus, the total in the linear shift register is $B_{MDAE}=l\cdot n_c+n_{IN}$ binary bits.

Through the linear shift register, a random bit from the original binary string is selected, and the entire binary string queue's head code is XORed with this bit, with the result entering the queue from the tail end. This step further compresses and encodes the data, generating a new binary string. After repeating this process B_{MDAE} times, the binary string in the linear shift register has been shuffled. A fixed-size sliding window scans the new binary string, and after each slide, the binary string within the window forms part of the new physical address. The design of the sliding window allows for more flexible and efficient data processing while also ensuring data continuity and integrity.

Assuming the window length is denoted by M_{SC}, and $M_{SC}>n_c$, if the window slides Y times after completing the scan, the following inequality gives the physical address length corresponding to o^k as $Y\cdot M_{SC}$:

$$Y \cdot M_{SC} < \beta \cdot l \cdot n_c, \beta \in (0,1) \tag{4}$$

In the context of image segmentation in complex backgrounds, the goal of this paper is to implement high-dimensional feature dimensionality reduction encoding through cerebellar neural networks to improve the efficiency and accuracy of image segmentation. Initially, it is necessary to determine the quantization levels of high-dimensional features. In this step, high-dimensional features are converted into forms that can be processed by machine learning models. This usually involves quantifying the features, i.e., converting continuous feature values into discrete level values. The choice of

quantization levels must consider the distribution of features and the complexity of the data, to preserve sufficient information while reducing computational complexity. Next, the receptive field scale of the virtual mapping encoding part is designed. In this step, based on the characteristics of the image and the requirements of the task, the scale of the receptive field is chosen to effectively capture and encode local information from the image. After completing the quantization and receptive field design, cerebellar neural networks can be used for feature dimensionality reduction encoding. Upon completing the feature dimensionality reduction encoding, it is necessary to determine the dimensions of the encoded features. In this step, based on the requirements of subsequent tasks and the capability of the model, the dimensions of the encoded features are selected to balance retaining sufficient information with reducing data and model complexity.

III. Image Data Augmentation in Complex Backgrounds

Image segmentation in complex backgrounds is a challenging task because complex backgrounds often contain various kinds of interference, such as noise, occlusions, and changes in lighting, all of which can impact the performance of image segmentation. Therefore, to enhance the accuracy and robustness of image segmentation, it is usually necessary to augment the image data before segmentation. This paper discusses the design of the image data augmentation module for complex background images, including the design of the generator and discriminator network structures and the specific implementation of self-play learning between the generator and discriminator, based on the concept of GAN.

In the task of image segmentation in complex backgrounds, the design of the generator network is the core part of the data augmentation module. Its task is to generate new image samples to enhance the dataset. This paper uses Gaussian noise as input to provide randomness, enabling the generator to produce a variety of images to ensure good generative performance across different datasets. Additionally, it is required that the dimension of Gaussian noise does not exceed the sample feature dimension to avoid the issue of dimensionality disaster. To ensure that the images output by the generator match the dimensionally reduced dataset, it is crucial that the output dimensions align with the dimensions of the samples after feature dimensionality reduction encoding. If the generator's output dimensions do not match the reduced sample dimensions, the generated images may not be correctly processed and utilized.

Assume the samples after feature dimensionality reduction encoding are represented by $z'=(z_1',z_2',...,z_p')$, and the Gaussian noise chosen by the generator network is represented by $b=(b_1,b_2,...,b_u,b_e)$, where $b_u \sim B(0,1)$, $e \leq o$, and the generator network has two hidden layers containing g_1 and g_2 neurons, respectively. Then, the generator network in the data augmentation module is a fully connected network of $e \times g_1 \times g_2 \times p$.

Let the weight matrices and bias vectors between two adjacent layers be represented by $(Q^{(1)}_{(g1\times e)},Q^{(2)}_{(g2\times g1)},Q^{(3)}_{(p\times g2)})$ and $(n^{(1)}_{(g1\times 1)},n^{(2)}_{(g2\times 1)},Q^{(3)}_{(p\times 1)})$, and the activation functions between the first hidden layer and the second hidden layer, and between the second hidden layer and the output layer are represented by $\delta^{(2)}$ and $\delta^{(3)}$, respectively. The input layer can be represented by $b=(b_1,b_2,...,b_u,...,b_e)$. The values of the neurons in the first hidden layer are $d^{(1)}=(d_1^{(1)},d_2^{(1)},...,d_k^{(1)},...,d_{g1}^{(1)})$, with an activation threshold $\phi^{(1)}$. The values of the neurons in the second hidden layer are $d^{(2)}=(d_1^{(2)},d_2^{(2)},...,d_k^{(2)},...,d_{g1}^{(2)})$, with an activation threshold $\phi^{(2)}$. The output layer outputs are represented by $z^{\sim'}=(z^{\sim'}_1,z^{\sim'}_2,...,z^{\sim'}_p)$, and the relationships between b, $d^{(1)}$, $d^{(2)}$, and $z^{\sim'}$ are given by the following equations:

$$d_k^{(1)} = \sum_{u=1}^{e} q_{ku}^{(1)} b_u + n_k^{(1)} \tag{5}$$

$$d_{8j}^{(2)} = \delta^{(2)}\left(\sum_{k=1}^{g_1} q_{jk}^{(2)} d_k^{(1)} + n_j^{(2)} - \varphi^{(1)} \right) \tag{6}$$

$$\tilde{z}_m' = \delta^{(3)}\left(\sum_{j=1}^{g_2} q_{mj}^{(3)} d_{8j}^{(2)} + n_m^{(3)} - \varphi^{(2)} \right) \tag{7}$$

In deep neural networks, the vanishing gradient problem is common, where the gradient values may become very small during backpropagation, affecting the model's learning and optimization. The use of *Leaky-ReLU* can prevent this issue, as it allows a small positive slope for negative input values, ensuring that gradients do not completely vanish even when inputs are negative. The expression for the *Leaky-ReLU* function is given by:

$$\delta(z) = \begin{cases} z, z > 0 \\ 0.01z, z \leq 0 \end{cases} \tag{8}$$

In the task of image segmentation in complex backgrounds, the main task of the discriminator network is to distinguish between images generated by the generator and real images. Therefore, its network structure needs to be capable of meeting this requirement. This paper employs a multi-layer Convolutional Neural Network (CNN) for the discriminator to extract higher-level and more abstract features, thereby enhancing its discriminatory ability.

In the practical scenario of image segmentation in complex backgrounds, the self-play learning between the generator and discriminator in the GAN is a dynamic, iterative process. In each iteration, the discriminator's parameters are fixed first to train the generator. Then, the discriminator judges the fake images produced, calculating a discrimination result. The goal of the generator is to maximize the probability of the discriminator making a mistake; thus, its loss function is typically the negative log probability of the discriminator's judgment of the fake images. Next, the generator's parameters are fixed to train the discriminator. The discriminator receives both the fake images produced by the generator and real images, and judges them. The goal of the discriminator is to correctly distinguish between real and fake images; therefore, its loss function is usually the sum of the log probability of the discrimination result of real images and the negative log probability of the discrimination result of fake images. The steps of the generation and discrimination phases are repeated, gradually updating the parameters of the generator and discriminator until the stopping criteria are met.

IV. Image Segmentation in Complex Backgrounds using Improved GANs

Image segmentation in complex backgrounds is an exceedingly challenging task due to the presence of various interference factors such as occlusions, changes in lighting, and color similarities. These factors can significantly impact the accuracy of image segmentation. Although traditional GANs can improve image segmentation performance to some extent, their generator and discriminator designs are typically simplistic and may not effectively handle the interferences present in complex backgrounds. Additionally, while increasing the depth of traditional GANs can enhance the model's expressive power, it also introduces issues such as gradient vanishing, overfitting, and increased computational cost. Moreover, deeper networks add to the complexity of the model, making training and optimization more challenging. Therefore, it is necessary to improve traditional GANs to maintain model performance while minimizing network depth as much as possible.

To enhance the model's ability to process high-frequency information and detailed textures, and to address difficulties in handling noise pollution and edge information extraction, this paper proposes incorporating multi-level wavelet channels, pixel attention modules, and edge enhancement modules into the traditional model. The setup of multi-level wavelet channels and pixel attention modules primarily addresses the recovery of high-frequency information and detailed textures. Wavelet transform, widely used in the time-frequency domain, allows for the analysis of images at various scales, accurately restoring high-frequency information. Meanwhile, pixel attention modules enable the model to focus on pixels that are more critical for the segmentation task, thereby better preserving intermediate features of the image, which is highly beneficial for the conservation of detailed textures.

Fig. 3 illustrates the basic working principle of the edge enhancement module. The setup of the edge enhancement module primarily addresses issues of noise pollution and edge information extraction. In complex backgrounds, image edges are often affected by various factors, such as noise pollution. The edge enhancement module utilizes mask operations to eliminate noise pollution while extracting and enhancing image edge contours, which is highly beneficial for improving the accuracy of image segmentation.

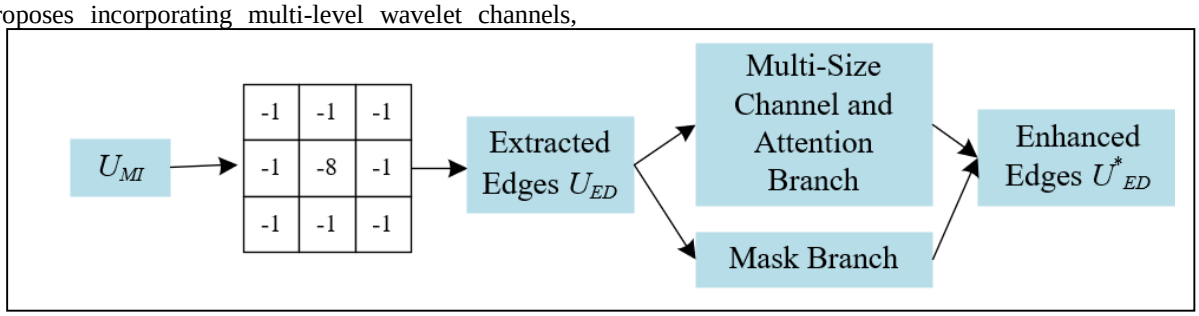


Fig. 3. Basic working principle of the edge enhancement module.

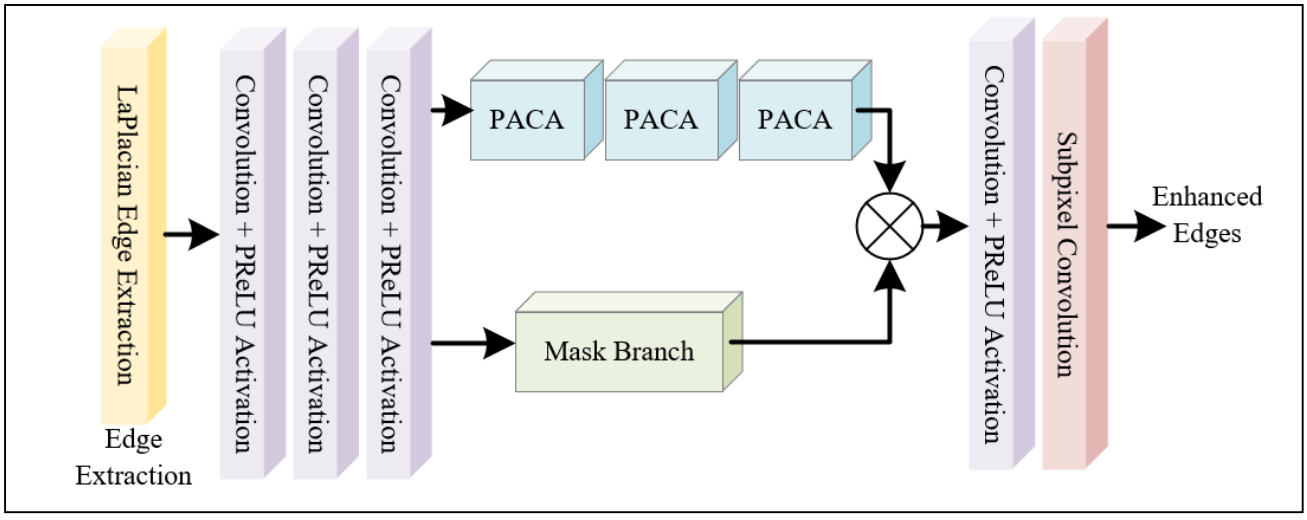


Fig. 4. Structure of the multi-scale channel and pixel attention network branches.

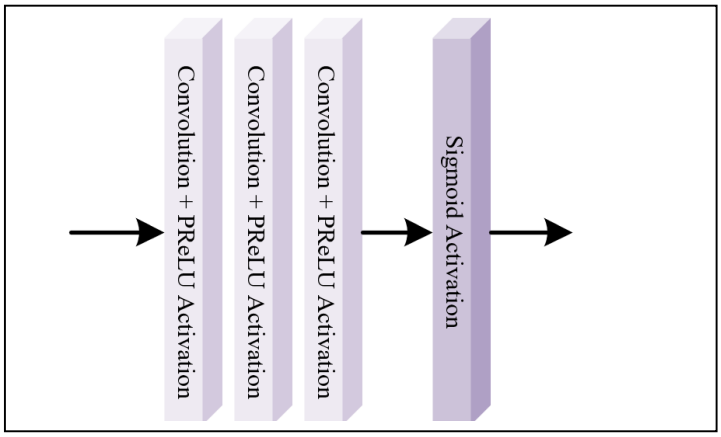


Fig. 5. The mask branch structure.

In the task of image segmentation in complex backgrounds, the extraction and utilization of edge information are crucial. Edge information typically helps to better distinguish between different target objects, leading to more accurate segmentation results. Therefore, the edge enhancement network in this paper includes two branches: the multi-scale channel and pixel attention Pixel Attention Channel Attention (PACA) branch, and the mask branch. Fig. 4 provides a schematic of the multi-scale channel and pixel attention network branch structure. The multi-scale channel and pixel attention PACA branch primarily handles the extraction of fine edge maps in images. In complex backgrounds, image edge information may be affected by noise and other factors, necessitating refined extraction through the PACA branch. The multi-scale channels allow for the analysis of images at various scales, which not only captures a wide range of background information but also detects fine detail. The pixel attention mechanism enables the model to focus on pixels that are crucial for edge extraction, thus better extracting fine edge maps. The role of the mask branch is to adaptively learn specific weight matrices and apply soft attention to relevant information. Fig. 5 provides a schematic of the mask branch structure. In processing complex background images, some irrelevant information such as noise and lighting changes may interfere. The mask branch can learn a weight matrix that filters out such irrelevant information, focusing only on those details that aid in edge extraction, thereby enhancing the accuracy of edge extraction.

Assuming the reconstructed intermediate image is represented by U_{MI}, edges derived from the intermediate image by U_{ED}, and enhanced image edges by U^*_{ED}. To ensure that the edge enhancement module inputs the intermediate image obtained from the multi-scale feature extraction module, this paper uses a Laplacian operator to mark edges on the intermediate SR image, thus better preserving edge information and improving segmentation accuracy. The Laplacian operator $M(z,t)$ for image $U(z,t)$ can be defined as its second derivative, as:

$$M(z,t)=\frac{\partial^2 U}{\partial z^2}+\frac{\partial^2 U}{\partial t^2} \tag{9}$$

Assuming the discrete convolution mask is represented by $M(z,t)$, the extracted edge map by $R(z,t)$, and the convolution operator by $\otimes$. The following equation describes the Laplacian process:

$$R(z,t)=M(z,t)\otimes U(z,t) \tag{10}$$

In actual images, the target might become unrecognizable due to scale changes, rotation, and other factors. Therefore, an effective method is needed to extract edge information that can also handle these issues. Strided convolution allows skipping some pixels in each convolution operation, ensuring computational efficiency to some extent and enabling the model to capture global information about the image. This is particularly helpful for image segmentation tasks in complex backgrounds. This paper utilizes changes in stride length of strided convolution to extract edge maps at different scales. Thus, even if the scale of the target changes, edge information can still be effectively extracted and further transformed into the LR space, combining high-resolution edge information with low-resolution background information, thus preserving fine edge details while capturing global background information, thereby improving segmentation accuracy. Assuming the downsampling operation of strided convolution is represented by $F(\cdot)$, the multi-scale channel and pixel attention network branch by $D(\cdot)$, the mask branch by $L(\cdot)$, and the upsampling operation of the subpixel convolution layer by $IO(\cdot)$, the operations in edge extraction can be characterized by the following equation:

$$U^*_{ED}=IO\left(D\left(F\left(U_{ED}\right)\right)\otimes L\left(F\left(U_{ED}\right)\right)\right) \tag{11}$$

Following these operations, the enhanced edge map is represented by U^*_{ED}.

In the task of image segmentation in complex backgrounds, segmentation based solely on features extracted by the model often has limitations, especially when facing complex and variable backgrounds, where the model may produce false positives or miss detections. To address this issue, this paper introduces adversarial loss and a discriminator. The task of the discriminator is to judge whether the generated image is close to a real image, while the adversarial loss measures the similarity between the generated and real images. Assuming the intermediate image similar to the HR image is represented by U_{BA}, a set of model parameters by $\phi_h=\{Q_{1:M};n_{1:M}\}$, the

Charbonnier penalty function by $\psi(z)=(z^2+\gamma^2)^{(1/2)}$, the real image after feature extraction by $U_{GE,s}$, and the features of the intermediate image after feature extraction by $U_{BA,u}$. The content loss function-based model used by the generator is represented by the following equation to generate U_{BA}:

$$M_V(\varphi_h) = \underset{\varphi_h}{\operatorname{argmin}} \sum_{s=1}^{b} \psi\left(U_{GE,s} - U_{MI,s}\right) \tag{12}$$

Through adversarial training, the discriminator compares the generated image with the real image, urging the generated image to be closer to the real image, thus improving the quality of segmentation. Assuming the model parameters in F_{NE} are represented by ϕ_f, the function of the generation network by $H(\cdot)$, and the discrimination function by $F(\cdot)$, the following equation provides the process expression for training the discriminator by minimizing the adversarial loss:

$$M_S(\varphi_f) = -\log F(U_{GE}) - \log\left(1 - F(H(U_{ME}))\right) \tag{13}$$

Charbonnier loss is a more general and robust loss function that performs well, especially in image reconstruction tasks such as image denoising and deblurring, focusing more on high-frequency details of images and providing better visual effects. To better reconstruct image details and reduce blur, this paper introduces pixel-based Charbonnier loss. Assuming the segmented image is represented by U_{AE} and the real image by U_{GE}, the paper introduces pixel-based Charbonnier loss to enhance the consistency of image content between the two. Assuming the model parameters in G are represented by ϕ_h. The real image and the final segmented image are represented by U_{GE} and U_{AE}, respectively, then the expression is:

$$M_{CST}(\varphi_h) = \psi(U_{GE} - U_{AE}) \tag{14}$$

Assuming the weight parameters for balancing loss components are represented by β and α, the final overall objective is given by the following equation:

$$M(\varphi_h, \varphi_f) = M_V(\varphi_h) + \beta M_S(\varphi_h, \varphi_f) + \alpha M_{CST}(\varphi_h) \tag{15}$$

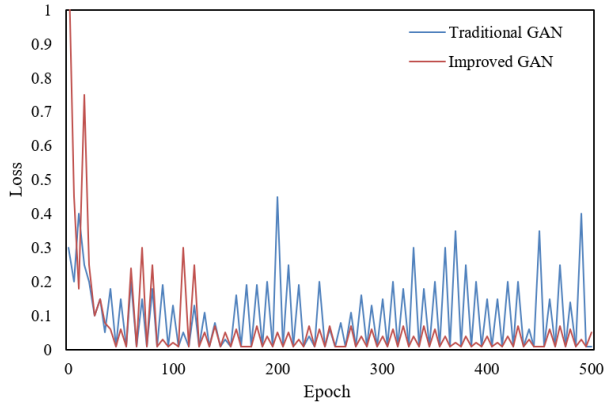
V. Experimental Results and Analysis

From the results shown in Table I, the model proposed in this paper achieved the best results across all four evaluation metrics: *R* (Recall), *F* (F-Value), *Dice* (Dice coefficient), and *mIoU* (mean Intersection over Union). Specifically, the proposed model reached a *Recall* of 0.9678, an *F-Value* of 0.9633, a *Dice* coefficient of 0.9631, and a *mIoU* of 0.9456. These values outperform other image segmentation models such as U-Net, Mask R-CNN, and DeepLab. The experimental results demonstrate the high effectiveness of the proposed method for image segmentation in complex backgrounds using improved GAN. Preprocessing of complex background image datasets and feature dimensionality reduction encoding using cerebellar neural networks, as well as image data enhancement in complex backgrounds, all contribute to enhancing the model's performance. Particularly, the newly designed generator and discriminator network structures, along with image data enhancement implemented through self-play learning, further improve the segmentation quality of our model.

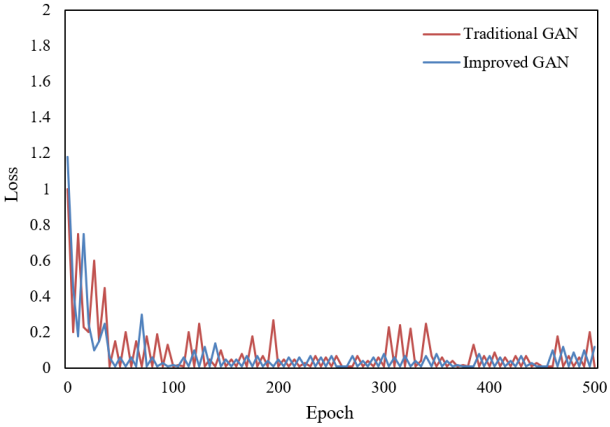
The image segmentation model for complex backgrounds developed in this paper employs both content loss and Charbonnier loss functions. From the loss function curves shown in the Fig. 6, the loss values of the improved GAN (red line) are lower than those of the traditional GAN (blue line) from the beginning of the training phase. As epochs increase, both models show a downward trend in loss values, indicating learning and optimization processes. However, the improved GAN maintains a lower overall level of loss, and the curve is relatively stable, demonstrating better stability and convergence throughout the training process. Especially in later epochs, the loss values of the improved GAN are significantly lower than those of the traditional GAN, suggesting that the improved GAN performs better in image reconstruction details and reducing blur. Moreover, the smaller fluctuations in the improved GAN's curve indicate strong robustness and relative stability during the training process.

TABLE I. Comparative Results of Different Image Segmentation Models

Method	R	F	Dice	mIoU
U-Net	0.8124	0.9127	0.9144	0.8124
Mask R-CNN	0.8365	0.9239	0.9236	0.8359
DeepLab	0.8257	0.9152	0.9127	0.8236
No Preprocessing	0.9456	0.9568	0.9568	0.9268
No Data Augmentation	0.9352	0.9211	0.9147	0.9347
The Proposed Model	0.9678	0.9633	0.9631	0.9456



(a) Content loss function



(b) Charbonnier loss function

Fig. 6. Loss function curves.

Combining both graphs, in the content loss function graph, the loss curve of the improved GAN decreases quickly early in training and then levels off, consistently maintaining a low level. In the Charbonnier loss function graph, although the improved GAN's loss curve fluctuates, the overall fluctuation amplitude is small, and the loss values consistently remain below those of the traditional GAN. This indicates that the improved GAN demonstrates better convergence and stability under different types of loss functions. It can be concluded that the improved GAN proposed in this paper outperforms traditional GANs in both content and Charbonnier loss functions. This not only manifests in lower loss values, potentially leading to higher quality image segmentation results, but also in training stability and convergence, where the improved GAN also shows superior performance. These results consistently and strongly validate the effectiveness of the proposed model, especially for image segmentation in complex backgrounds.

Fig. 7 shows the image segmentation accuracy at different iteration counts. It is evident that as the number of iterations increases, the accuracy at each epoch improves to varying degrees. From epoch 50 to epoch 150, there is a significant increase in accuracy across all iteration counts, particularly rapid in the early stages. For example, at 150 iterations, accuracy improved from 0.76 at epoch 50 to 0.94 at epoch 150, an increase of 18 percentage points. As training progresses, the rate of accuracy improvement begins to slow, but it still shows an upward trend. For instance, from iteration 200 to iteration 350, the later epochs do not see as rapid an increase as the initial phases, but with increasing epochs, the final accuracy continues to steadily improve. Higher iteration counts do not always significantly impact final accuracy. In some cases, additional iterations may lead to minor improvements or even stabilize. For example, at 350 iterations, from epoch 300 to epoch 350, the accuracy only improved by 0.01. Fig. 7 demonstrates that as the model training iterations increase, the accuracy of image segmentation generally improves, especially noticeable in the early training phases, highlighting the effectiveness of the model proposed in this paper. As iterations increase, accuracy continues to improve steadily, indicating that the model can effectively learn and adapt to the image segmentation task in complex backgrounds.

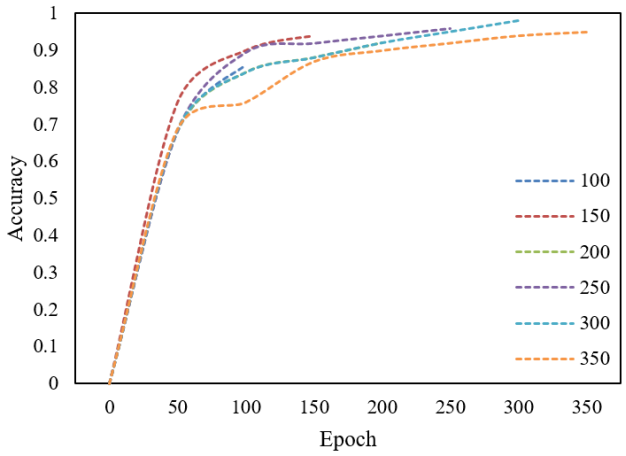


Fig. 7. Impact of iteration count on accuracy.

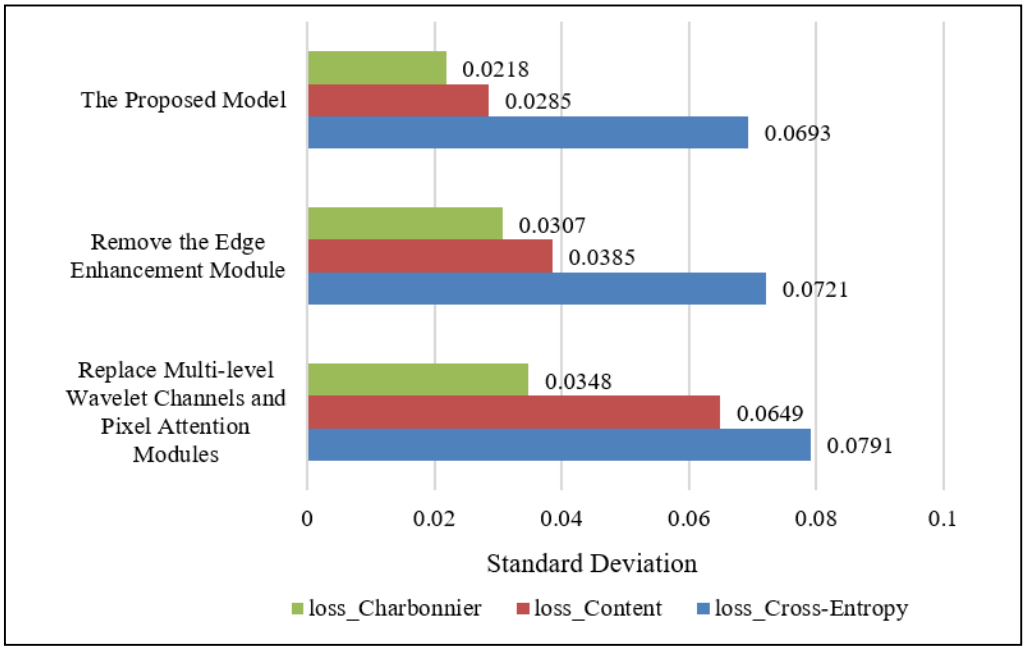


Fig. 8. Standard deviation of loss functions.

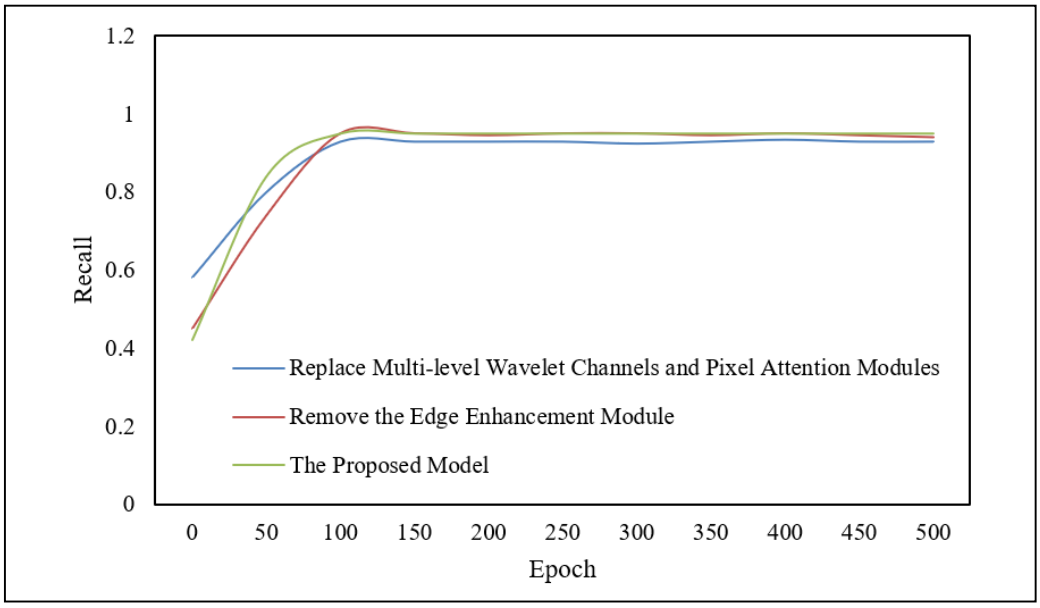


Fig. 9. Recall rate curve.

In Fig. 8, the proposed model and two variant models use three different loss functions: cross-entropy loss, content loss, and Charbonnier loss. The standard deviation of the loss functions can indicate the model's stability during training. Lower standard deviation values generally mean that the loss functions fluctuate less during training, indicating a more stable model. The proposed model shows lower standard deviation across all three loss functions, especially exhibiting the lowest values with Charbonnier loss, indicating high training stability. Models without the edge enhancement module show increased standard deviations across all loss functions, particularly with cross-entropy and content loss, suggesting that the edge enhancement module helps maintain stability during training. Replacing the multi-level wavelet channels and pixel attention modules leads to further increased standard deviations in all three loss functions, most notably in cross-entropy loss, demonstrating the importance of these modules in reducing fluctuations and enhancing model stability during training. The results show that the proposed model has the lowest standard deviation under various loss functions, indicating the strongest stability. This emphasizes the effectiveness of the proposed model in image segmentation tasks in complex backgrounds compared to models lacking key modules. Particularly, the edge enhancement module and multi-level wavelet channels and pixel attention modules prove to be indispensable parts of the model, contributing to improved stability during training and ultimate performance in image segmentation.

The three curves in Fig. 9 display the recall rates of three models at different training stages (epochs). It can be observed that in the early stages of training, all models quickly improve their recall rates as epochs increase, indicating that the models rapidly learn from the data and enhance their segmentation performance. As the training progresses, the recall rates of all models tend to stabilize, with the proposed model (green curve) showing quicker stability and achieving a higher final recall rate, suggesting that the model can stably identify true positive cases and cover them more comprehensively. Overall, the proposed model consistently outperforms the other two variant models in recall rate at nearly all training stages. The models without the edge enhancement module (red curve) and those replacing the multi-level wavelet channels and pixel attention module (blue curve) have similar recall rates but are lower than the proposed model. The above figure reflects the evident effectiveness of the proposed model for image segmentation in complex backgrounds, maintaining a high recall rate across multiple training stages. Particularly with the introduction of the edge enhancement module and multi-level wavelet channels and pixel attention modules, the model exhibits a higher recall rate, highlighting the significant contribution of these modules to improving image segmentation performance. In summary, the proposed model demonstrates good recall rates in the task of image segmentation in complex backgrounds, validating its design's rationality and efficiency.

VI. Conclusion

This paper addressed the challenging task of image segmentation in complex backgrounds, where interference factors are difficult to distinguish and highly variable. The research includes preprocessing complex background image datasets and using cerebellar neural network-based methods for feature dimensionality reduction encoding to more effectively handle high-dimensional data. The paper also focuses on designing data enhancement strategies for complex background images to generate more diverse training data and improve the model's generalizability. An improved GAN was proposed, featuring new generator and discriminator network structures optimized through self-play learning to enhance data augmentation effects.

Experimental comparisons of traditional GAN, improved GAN, and variant GAN models without key modules showed that the improved GAN displays better performance and higher stability in terms of loss function reduction speed, final stable values, and standard deviation during training. The proposed model achieved a recall (R) of 0.9678, an F-score (F) of 0.9633,

a Dice coefficient of 0.9631, and a mean Intersection over Union (mIoU) of 0.9456. These metrics surpass those of other image segmentation models, such as U-Net, Mask R-CNN, and DeepLab. In terms of recall rate, the improved GAN demonstrates quicker improvement and a higher final recall rate compared to other variants, proving its effectiveness. The introduced Charbonnier loss function and edge enhancement module both play positive roles in improving the model's segmentation accuracy and robustness.

The improved GAN proposed in this paper shows outstanding performance in image segmentation tasks in complex backgrounds. By combining cerebellar neural network-based feature dimensionality reduction encoding and innovative generator and discriminator network structures, the model's performance has been significantly enhanced. The network, after self-play learning and data enhancement, can generate high-quality segmentation results, maintaining high accuracy and stability even in complex backgrounds. A series of experiments have validated the model's effectiveness, with its ability to handle high-frequency information, edge contours, and detailed textures surpassing traditional methods, making it highly potential for practical applications.

REFERENCES

[1] Z. Zhang, Y. Tian, and J. Zhang, "Complex image background segmentation for cable force estimation of urban bridges with drone-captured video and deep learning," Struct. Control Health Monit., vol. 29, no. 4, e2910, 2022. https://doi.org/10.1002/stc.2910.

[2] U. Tatli and C. Budak, “Biomedical image segmentation with modified U-Net,” Trait. Signal., vol. 40, no. 2, pp. 523-531, 2023. https://doi.org/10.18280/ts.400211.

[3] Z. Chen, C. Li, Z. Jiang, and Y. Zhao, "Application of fuzzy C-means algorithm in complex background image segmentation of forensic science," in Commun. Signal Process. Syst. - Proc. 2018 CSPS, vol. 517, 2020, pp. 212-217. https://doi.org/10.1007/978-981-13-6508-9_27.

[4] M. Aggarwal, A.K. Tiwari, and M.P. Sarathi, “Comparative analysis of deep learning models on brain tumor segmentation datasets: BraTS 2015-2020 datasets,” Rev. Intell. Artif., vol. 36, no. 6, pp. 863-871, 2022. https://doi.org/10.18280/ria.360606.

[5] Y. Ma, Y. Zhang, and S. Yang, "Soil image segmentation algorithm under complex background," J. Phys.: Conf. Ser., vol. 1549, no. 2, 022036, 2020. https://doi.org/10.1088/1742-6596/1549/2/022036.

[6] A. Rehman, M. A. Butt, and M. Zaman, “Liver Lesion Segmentation Using Deep Learning Models,” Acadlore Trans. Mach. Learn., vol. 1, no. 1, pp. 61-67, 2022. https://doi.org/10.56578/ataiml010108.

[7] S. Srivastava, G. Kumar, R.K. Mishra, and N. Kulshrestha, “A complex diffusion based modified fuzzy C- means approach for segmentation of ultrasound image in presence of speckle noise for breast cancer detection,” Rev. Intell. Artif., vol. 34, no. 4, pp. 419-427, 2020. https://doi.org/10.18280/ria.340406.

[8] V. Gavini and G.R.J. Lakshmi, “CT image denoising model using image segmentation for image quality enhancement for liver tumor detection using CNN,” Trait. Signal., vol. 39, no. 5, pp. 1807-1814, 2022. https://doi.org/10.18280/ts.390540.

[9] A.H. Alwan, S.A. Ali, and A.T. Hashim, “Medical image segmentation using enhanced residual U-Net architecture,” Math. Model. Eng. Problems., vol. 11, no. 2, pp. 507-516, 2024. https://doi.org/10.18280/mmep.110223.

[10] R. Sille, T. Choudhury, P. Chauhan, and D. Sharma, “Dense hierarchical CNN – A unified approach for brain tumor segmentation,” Rev. Intell. Artif., vol. 35, no. 3, pp. 223-233, 2021. https://doi.org/10.18280/ria.350306.

[11] R. Sille, T. Choudhury, P. Chauhan, and D. Sharma, “A systematic approach for deep learning based brain tumor segmentation,” Ing. Syst. Inf., vol. 26, no. 3, pp. 245-254, 2021. https://doi.org/10.18280/isi.260301.

[12] S. Z. Rahman, T. R. Singasani, and K. S. Shaik, “Segmentation and Classification of Skin Cancer in Dermoscopy Images Using SAM-Based Deep Belief Networks,” Healthcraft Front., vol. 1, no. 1, pp. 15-32, 2023. https://doi.org/10.56578/hf010102.

[13] V. R. S. R. Nagireddy and K. S. Shaik, “Advanced Hybrid Segmentation Model Leveraging AlexNet Architecture for Enhanced Liver Cancer Detection,” Acadlore Trans. Mach. Learn., vol. 2, no. 3, pp. 116-128, 2023. https://doi.org/10.56578/ataiml020301.

[14] B. Xie, Y. Heng, F. Shuyi, Y. Wang, and X. Zhu, "Sea-land segmentation of infrared remote sensing image based on complex background," in Proc. Int. Conf. Comput. Vis. Image Deep Learn. (CVIDL), Chongqing, China, 2020, pp. 165-168. https://doi.org/10.1109/CVIDL51233.2020.00039.

[15] S. Sacharisa and I.H. Kartowisastro, “Enhanced spine segmentation in scoliosis X-ray images via U-Net,” Ing. Syst. Inf., vol. 28, no. 4, pp. 1073-1079, 2023. https://doi.org/10.18280/isi.280427.

[16] J. Huang, Q. Yuan, D. Liu, and H. Fu, "An image segmentation method for banana leaf disease image with complex background," in Proc. 5th Int. Conf. Data Sci. Inf. Technol. (DSIT), Shanghai, China, 2022, pp. 1-5. https://doi.org/10.1109/DSIT55514.2022.9943850.

[17] A. Abo-El-Rejal, S. E. Ayman, and A. Aymen, “Advances in Breast Cancer Segmentation: A Comprehensive Review,” Acadlore Trans. Mach. Learn., vol. 3, no. 2, pp. 70-83, 2024. https://doi.org/10.56578/ataiml030201.

[18] J. L. Cui, B. Lian, G.-Y. Kang, Z. M. Lu, and Y. C. Chen, "Image segmentation and center positioning method for roughcast wheel hubs under complex background," J. Netw. Intell., vol. 6, no. 1, pp. 54–63, 2021.

[19] R. Yang, X. J. Qian, and B. B. Zhang, "Multi-feature fusion aerial image segmentation in complex background," in Proc. 3rd Int. Conf. Vis. Image Signal Process. (ICVISP), Vancouver BC, Canada, 2019, pp. 1-8. https://doi.org/10.1145/3387168.3387237.

[20] K. Li, Q. Feng, and J. Zhang, "Co-segmentation algorithm for complex background image of cotton seedling leaves," J. Comput.-Aided Des. Comput. Graph., vol. 29, no. 10, pp. 1871-1880, 2017.

[21] A. M. Abdeldaim, E. H. Houssein, and A. E. Hassanien, “Color image segmentation of fishes with complex background in water,” in Advances in Intelligent Systems and Computing, vol. 723, pp. 634–643, 2018. https://doi.org/10.1007/978-3-319-74690-6_62.

[22] M. Barthakur, K. K. Sarma, and N. Mastorakis, “Learning aided structures for image segmentation in complex background,” in Proc. 2017 Eur. Conf. Electr. Eng. Comp. Sci. (EECS), Bern, Switzerland, 2017, pp. 158–165. https://doi.org/10.1109/EECS.2017.39.

[23] L. Liu, X. Cheng, J. Dai, and J. Lai, “Adaptive threshold segmentation for cotton canopy image in complex background based on logistic regression algorithm,” Trans. Chinese Soc. Agric. Eng., vol. 33, no. 12, pp. 201-208, 2017.

[24] M. Yuan, Y. Li, J. Sun, B. Shi, J. Xu, L. Xu, and Y. Wang, “Distributed Learning based on Asynchronized Discriminator GAN for remote sensing image segmentation,” in Proc. 8th Int. Conf. Commun. Inf. Process., Beijing, China, 2022, pp. 33-40. https://doi.org/10.1145/3571662.3571668.

[25] X. Feng, J. Lin, C. M. Feng, and G. Lu, “GAN inversion-based semi-supervised learning for medical image segmentation,” Biomed. Signal Process. Control, vol. 88, 105536, 2024. https://doi.org/10.1016/j.bspc.2023.105536.

[26] K. Zhang, Y. Shi, C. Hu, and H. Yu, “Nucleus image segmentation method based on GAN and FCN model,” Soft Comput., vol. 26, no. 16, pp. 7449-7460, 2022. https://doi.org/10.1007/s00500-021-06449-y.

[27] C. J. Shin and Y. S. Heo, “GAN Inversion with Semantic Segmentation Map for Image Editing,” in 2022 13th Int. Conf. Inf. Commun. Technol. Convergence (ICTC), Jeju Island, Korea, 2022, pp. 927-931. IEEE. https://doi.org/10.1109/ICTC55196.2022.9952548.

[28] H. X. Wang and N. Liu, “Research on fingerprint image segmentation under complex background,” Appl. Mech. Mater., vols. 687-691, pp. 3612-3615, 2014. https://doi.org/10.4028/www.scientific.net/AMM.687-691.3612.

[29] M. Li and M. Bai, "Text segmentation from image with complex background based on Markov random fields," in Proc. 2012 Int. Conf. Computer Science and Electronics Engineering, ICCSEE, Hangzhou, China, 2012, pp. 486–489. https://doi.org/10.1109/ICCSEE.2012.404.

[30] H. Hong, X. Guo, and X. Zhang, "An improved segmentation algorithm of color image in complex background based on graph cuts," in Proc. 2011 IEEE Int. Conf. on Computer Science and Automation Engineering, CSAE 2011, Shanghai, China, 2011, pp. 642–645. https://doi.org/10.1109/CSAE.2011.5952551.

[31] Z. Diao, H. Wang, Y. Song, and Y. Wang, "Segmentation method for cotton mite disease image under complex background," Transactions of the Chinese Society of Agricultural Engineering, vol. 29, no. 5, pp. 147–152, 2013.

[32] A. Beji, A. G. Blaiech, M. Said, A. B. Abdallah, and M. H. Bedoui, "An innovative medical image synthesis based on dual GAN deep neural networks for improved segmentation quality," Applied Intelligence, vol. 53, no. 3, pp. 3381–3397, 2023. https://doi.org/10.1007/s10489-022-03682-2.

[33] D. Zhai, B. Hu, X. Gong, H. Zou, and J. Luo, "ASS-GAN: Asymmetric semi-supervised GAN for breast ultrasound image segmentation," Neurocomputing, vol. 493, pp. 204–216, 2022. https://doi.org/10.1016/j.neucom.2022.04.021.

[34] L. Zhao, "Image semantic segmentation method based on GAN network and FCN model," Journal of Engineering, vol. 2022, no. 1, pp. 1–9, 2022. https://doi.org/10.1049/tje2.12085.

[35] M. L. Dimoiu, D. Popescu, and L. Ichim, "Improved conditional GAN for aerial image segmentation," in 2021 IEEE AFRICON Conference, AFRICON, Arusha, Tanzania, 2021, pp. 1–6. https://doi.org/10.1109/AFRICON51333.2021.9570942.